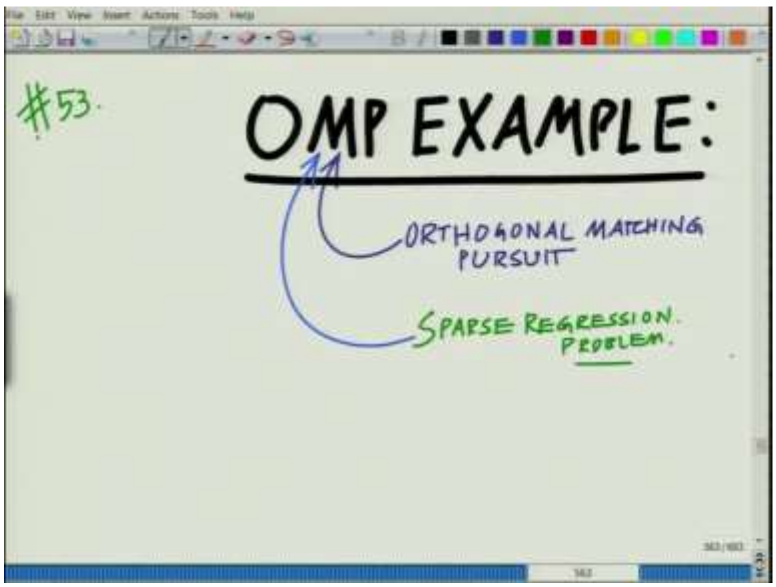
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor. Aaditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture No. 53
OMP Example for Sparse Regression

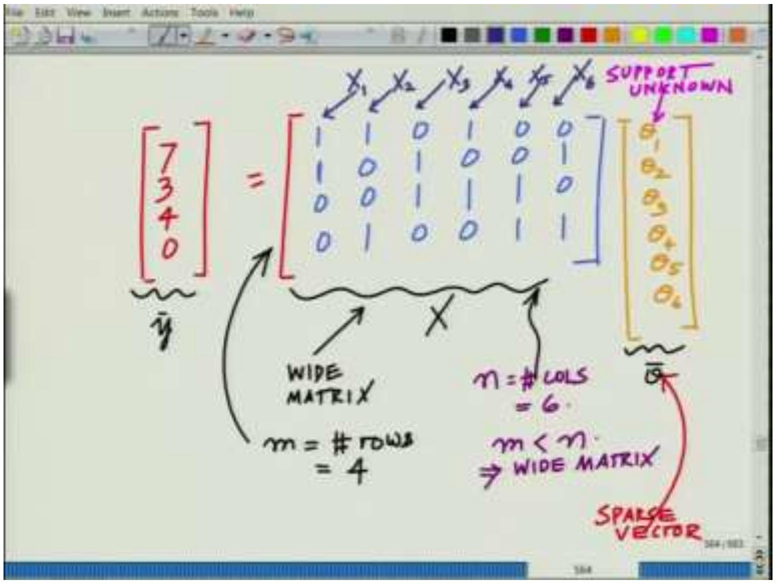
Hello! Welcome to another module in this massive open online course. So, we are looking at sparse regression, let us continue our discussion by example. We have looked at the OMP algorithm that is the Orthogonal Matching Pursuit to solve this sparse regression problem. So, let us now look at a specific example involving the OMP to understand the procedure better.

(Refer Slide Time: 00:38)



So, we have seen the OMP, now let us look at an example to understand this better. Remember OMP stands for Orthogonal Matching Pursuit and this is an algorithm for the sparse regression problem, algorithm to solve this sparse regression problem.

(Refer Slide Time: 01:44)



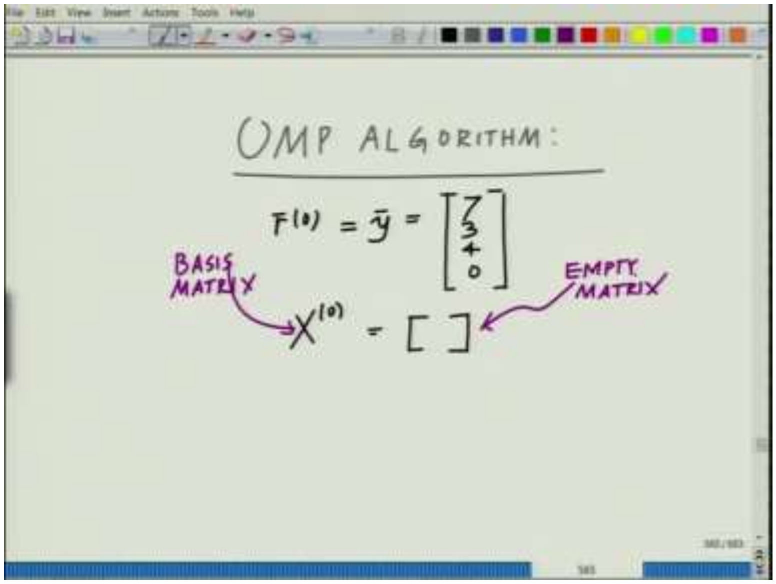
So, let us look at a simple example. So, the example I want to consider for sparse regression is the following thing. So, we have the vector y bar which is given by this which is equal to, and we have the vector that is, so this is your vector and let us denote this by y bar. Let us denote this matrix by X. This we denote by theta bar and in fact if you can look at this, this is what we are calling as a wide matrix.

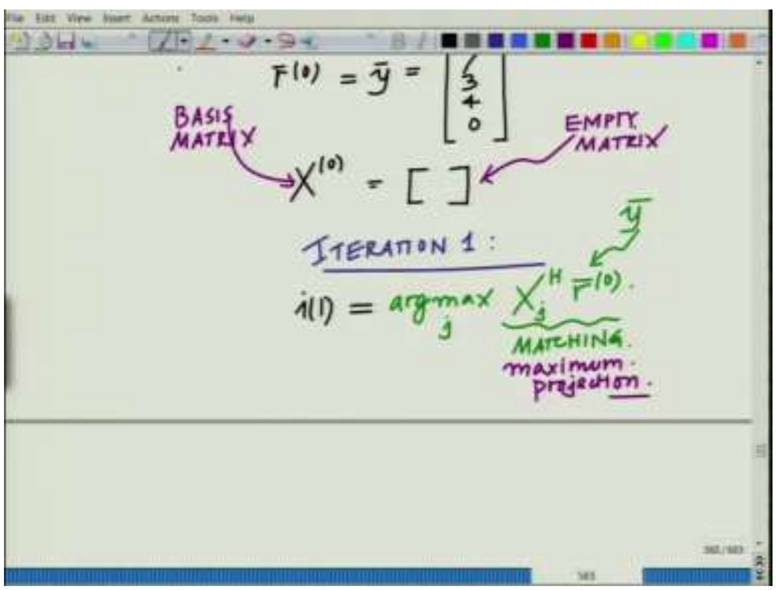
You can see because here you have m which is equal to the number of rows and this is equal to 4 and you have n which is equal to the number of columns, this is 6. So, we have m less than n. So, this implies this becomes a wide matrix because the number of measurement is less than the number of unknowns and this theta, remember this theta, this is a, we have to find the theta such that this theta is a sparse vector.

Remember, many of the components of theta bar that is many theta i are 0. Only very few theta i are non 0 but the catch is we do not know which theta i are non 0. Now, if we knew which theta i are non 0, then we would simply retain those theta i and the columns in X corresponding to those theta i and solve the problem. So, that would become an easy problem to solve. Unfortunately what makes this complex, the problem rather challenging that it is not known which theta i are 0 and which theta i, that is the support of this vector.

The support of the vector indicates those indices where you have the non 0 theta i, the support of this parameter vector theta bar is unknown. So, while say simply here that you have the support that is the key in sparse regression that is the support is unknown and these are the columns, remember we said X1, X2, X3, X4, X6. These are the columns and now we want start the OMP procedure which proceeds as follows.

(Refer Slide Time: 05:28)



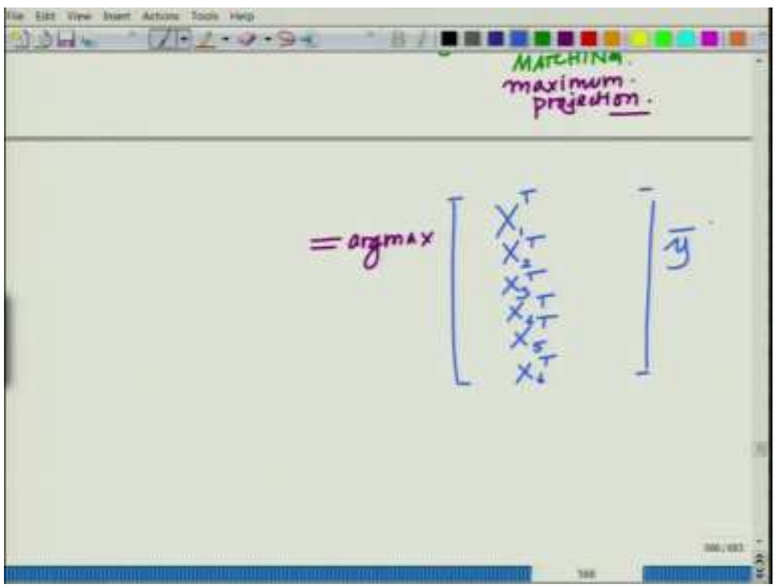


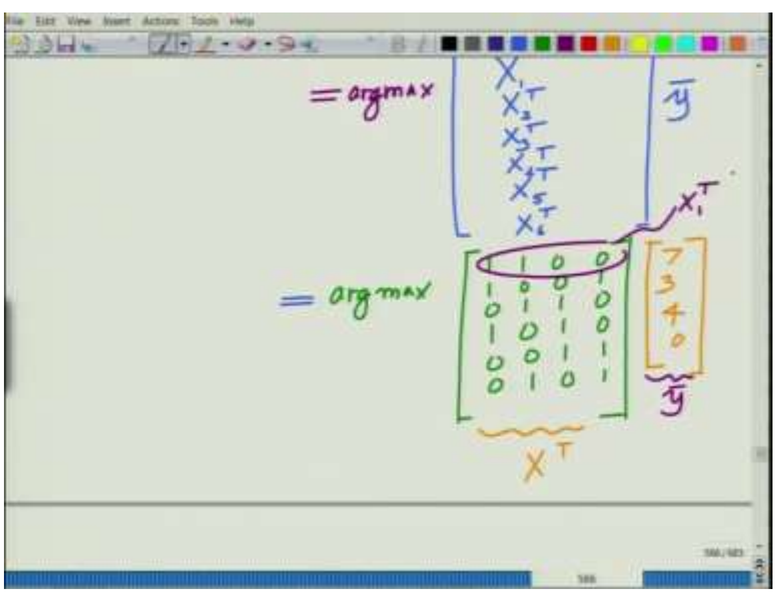
So, we want to start the OMP algorithm, or let us call it as the OMP algorithm. So, you have the OMP algorithm and what happens in the OMP algorithm? Remember we start with this notion of residue. So, we set the residue 0. This is equal to y bar which is basically equal to 7, 3, 4, 0 and we set X of 0. Remember, this is basically your, a set. Remember X, this is the basis matrix or the, what we use for subset selection, selecting the subsets of the columns of the matrix X i.

Initially, this is initialized as the empty matrix to begin with. Now, we begin with the iteration 1. The first iteration. So, now what happens, now once we have defined these two quantities one can start with the first iteration. So, what is the first iteration? Remember, we have to find the index that satisfies, that has the maximum projection, that is the column of X on which the y bar or essentially the residue at from the previous iteration, which is in this case is nothing but y bar has the maximum projection.

So, this is basically your, remember this is the matching step in the matching pursuit or essentially you are trying to find the column which has the maximum projection.

(Refer Slide Time: 07:56)

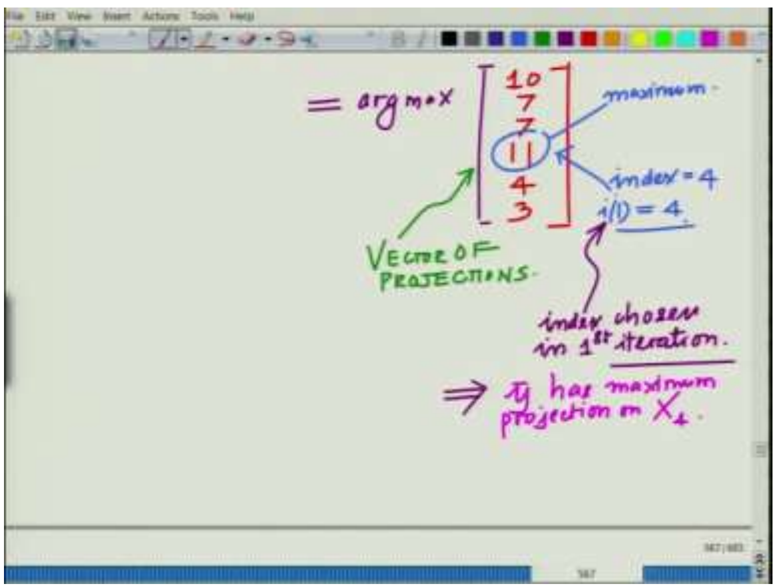




Now, another way to do it is I can simply look at X bar and take X transpose because X transpose will essentially contain the columns that is X1 transpose, X2 transpose, X3 transpose, X4 transpose, X5 transpose and X6 transpose times, so I will just, I can simply take the product X transpose times y bar because this will give me all the projections.

So, this I can write this essentially I take the matrix 1 1 0 0, 1 0 0 1, 0 1 1 0, 1 0 1 0, 0 0 1 1 and 0 1 0 1. This is essentially your matrix. This is the matrix that I have times y bar which is 7, 3, 4, 0. So, this is basically your X transpose, this is basically y bar. So, this you can see essentially, this is basically your X 1 transpose and so on and therefore what this is going to give me?

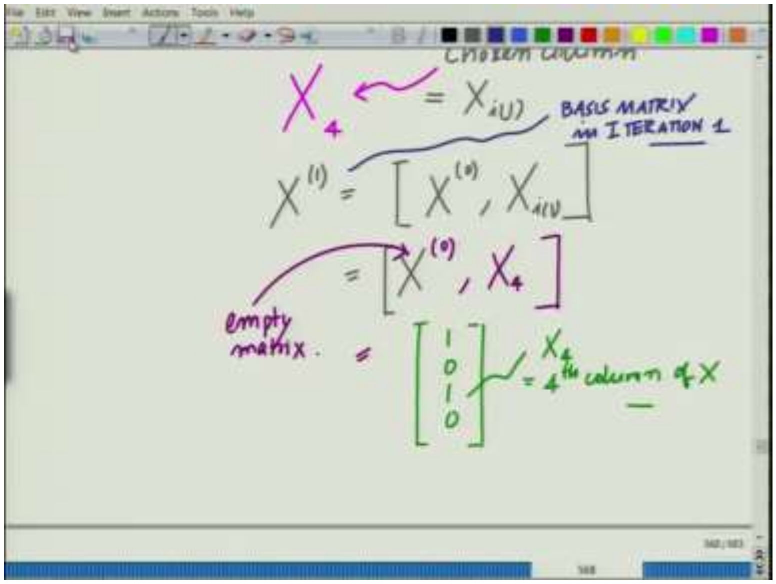
(Refer Slide Time: 09:50)



I am going to take the maximum of these projections and you can evaluate these projections. The first one will be 1 1 0 0 times 7 3 4 0 projection 7 3 4 0 so that will be 7 plus 3. The projection will be 10. Similarly, the next one would be 7 plus 0 that is 7, then the next one, fourth one would be 7 plus 4 that is 11, then fifth one would be 4 and the final one would 3 plus 0 that is 3. So, this is essentially the vector of projections.

Remember vector projections of y in the columns of X. So, this is the vector of, and in this we take the maximum, and you can see this is basically the maximum but we are not interested in the maximum value. We are interested in the index of this maximum value. So, i 1, the index remember is equal to 4, so this implies the index chosen in the first iteration that is i 1 equal to 4. So, this is basically the index chosen in the 5 first iteration. This is the index chosen in the first iteration, there is i 4. The i 1 equal to 4. This implies X4 has the maximum projection or implies y bar has maximum projection on X4.

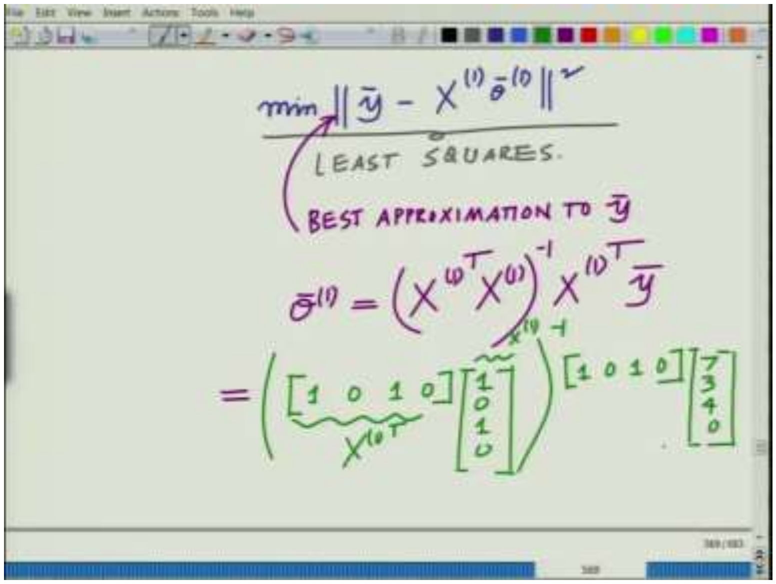
(Refer Slide Time: 12:16)



And therefore, now our chosen columns, so 4 this is basically which is essentially equal to X of i 1, this becomes your chosen column and X 1 is basically obtained by remember augmenting X0 with Xi 1 that is basically augmenting X0 with X4. Now, this is the empty matrix. So, this simply becomes equal to X4 which is essentially the column, if you recall that is the fourth column, that is 1 0 1 0. This is basically X4 which is fourth column of X and this is essentially the basis matrix that is constructed in the first iteration.

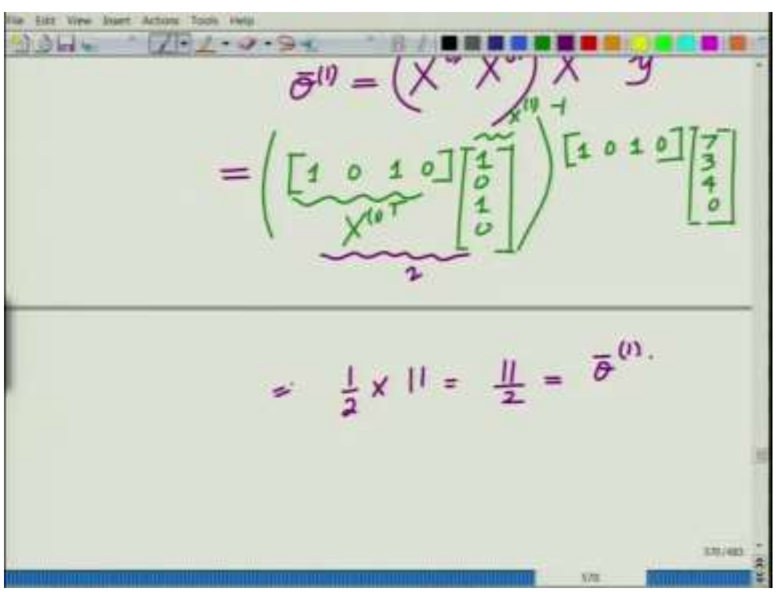
So, this is the basis matrix and now we have to find the best approximation to y bar. Remember now, we have to solve the least squares problem.

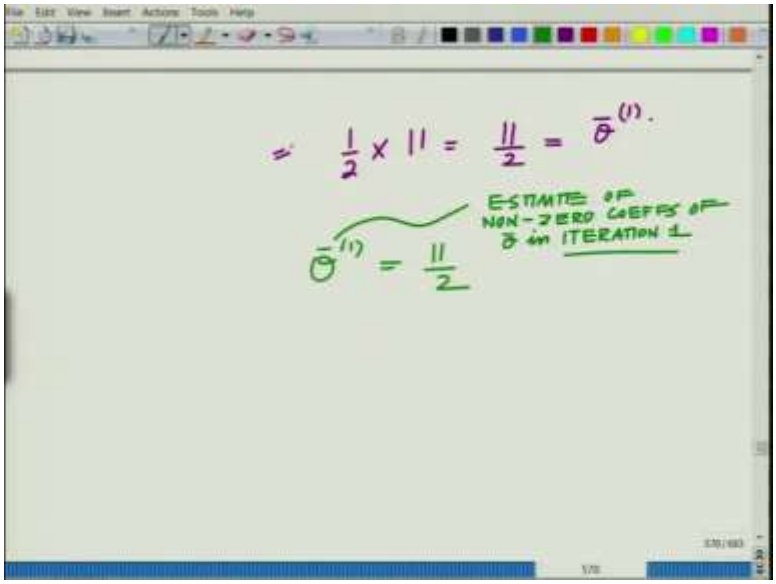
(Refer Slide Time: 14:04)



So, now we are using this matrix, we do y bar minus remember X of 0, X of 1 theta bar of 1. So, what we are doing is this is? Essentially our least squares problem and we are finding the best approximation to y bar and the solution is given as the pseudo inverse of X that is X1 transpose X1 inverse times X1 transpose y bar which essentially, now if you substitute X1 that is nothing but you have the row vector 1 0 1 0 times 1 0 1 0. This is basically your X1 transpose and this is your X1. So, inverse X1 transpose which is 1 0 1 0 times your y bar which is 7 3 4 0.

(Refer Slide Time: 15:55)

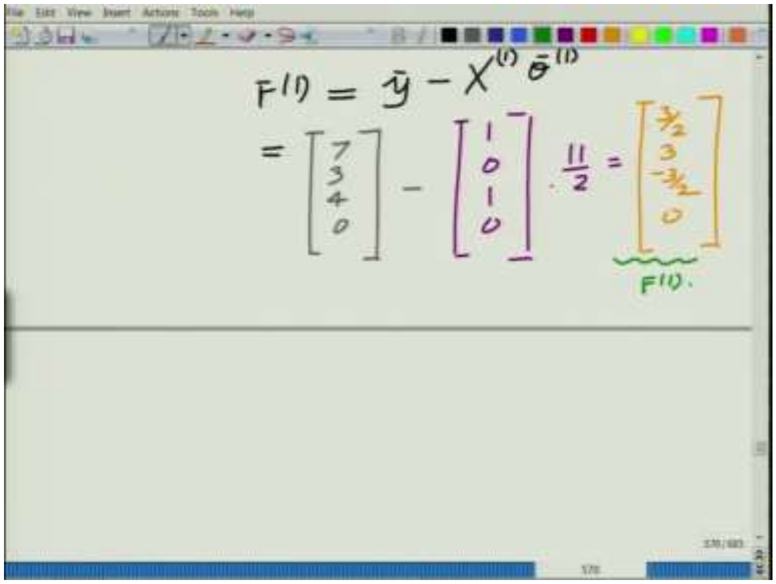


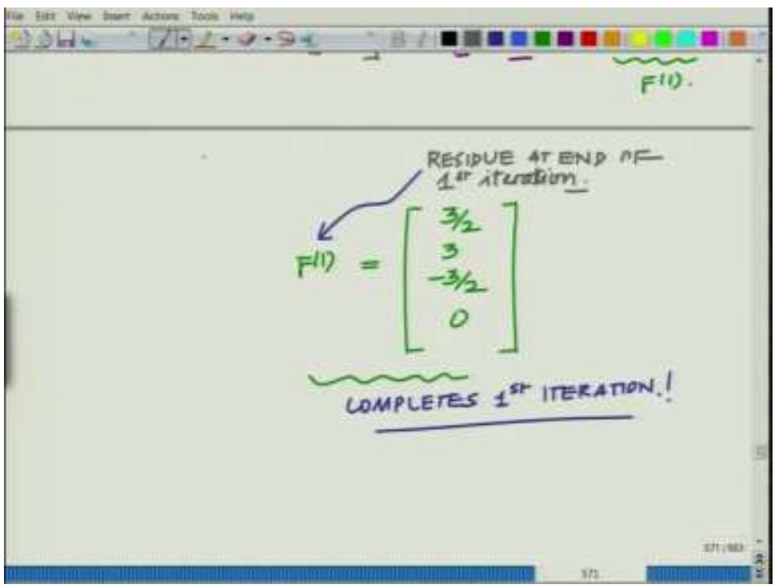


And this you can see the inside of this that X 1 transpose X 1 this evaluates to 2 so inverse of that this is equal to half times 1 0 1 0 times 7 3 4 0 so that is 7 plus 4, 11 so this becomes equal to 11 by 2. So, this is basically your theta bar 1. So, the estimate, so the estimate of theta bar, in fact I would say estimate of, theta bar in the first iteration. This is the estimate of the non 0 coefficients because remember we are estimating the non 0 coefficients.

We are trying to find the matching columns and basically assuming that these represent the locations of, because remember you have to construct an approximation y bar with the fewest column of the X. So, you are progressively isolating those columns of X which you think are weighted by non 0 values of theta in the linear combination.

(Refer Slide Time: 17:23)

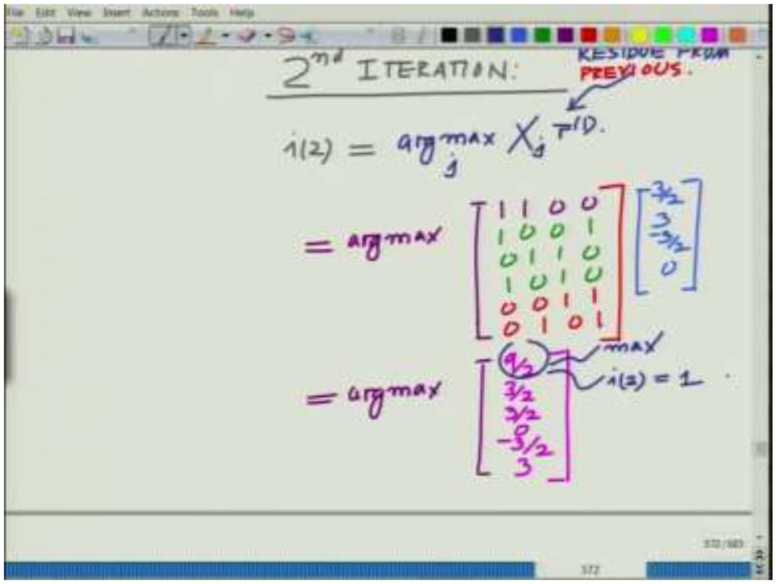


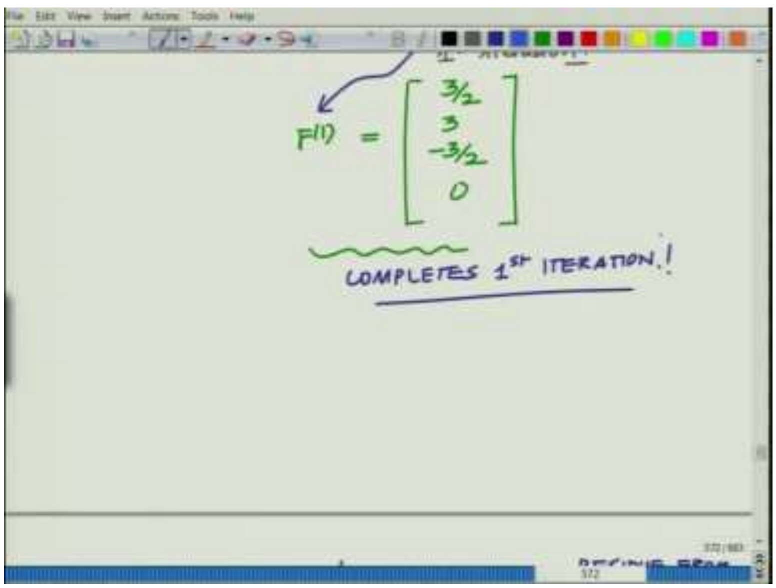


Now that brings us to r, we have to compute the residue to complete this iteration so the residue is basically y bar minus X 1 times theta bar 1 so if you substitute this that becomes 7 3 4 0 minus X 1 is basically 1 0 1 0 times 11 by 2 which is equal to, if you know write this, this will be equal to, you can clearly see this is 3 by 2, 7 minus 11 by 2, 3 by 2, 3 minus 0. This is 3, 4 minus 11 by 2 minus 3 by 2 and 0, that is essentially what you are going to have and this is basically your residue at the end of the first iteration.

So, your r bar 1 that is basically equal to 3 by 2, 3 minus 3 by 2. So, this is basically completes the first iteration and this is your residue at the end of the first iteration.

(Refer Slide Time: 19:37)





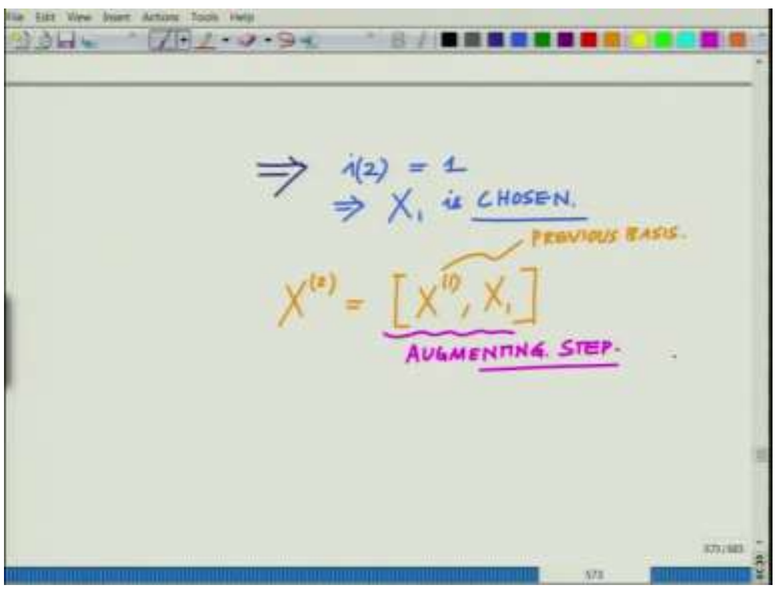
Now the point is, now we want to perform, now we begin the second iteration and second iteration will be very simple. It is a repetition of all the steps again. In order what we have done in the first iteration, that is why it is called an iteration, iteration essentially means a set of steps that you keep repeating again and again until you achieve the desired solution.

So, now the only difference is the, you will now project on the residue from the previous iteration which is r bar 1. So, essentially now the projection step will be that is i 2 this will be your argmax j project on r bar 1. This is residue, this is the only change and then you will repeat residue from previous iteration, and this will be again once again, let us compute using the transpose of the matrix. So, this will be once again remember 1 1 0 0, 1 0 0 1, 0 1 1 0, 1 0 1 0, 0 0 1 1 and 0 1 0 1.

And now I will have r bar of, I will use r bar of 1. And we know r bar of 1 is essentially what we have derived over here 3 by 2, 3 minus 3 by 2, 0, and you do the projection, you can evaluate these projections, you will find that these values are given by 3 by 2 plus 3 by 2 that is 9 by 2, 3 by 2 plus 0, 3 by 2 minus 3 by 2 that is 3 by 2. Then you have 3 by 2 minus 3 by 2, 0, 0 0 1 1 minus 3 by 2 and 0 1 0 1 so this is 3 and here now you can clearly see that this is the maximum but once again we want only the index. So, the index is 1.

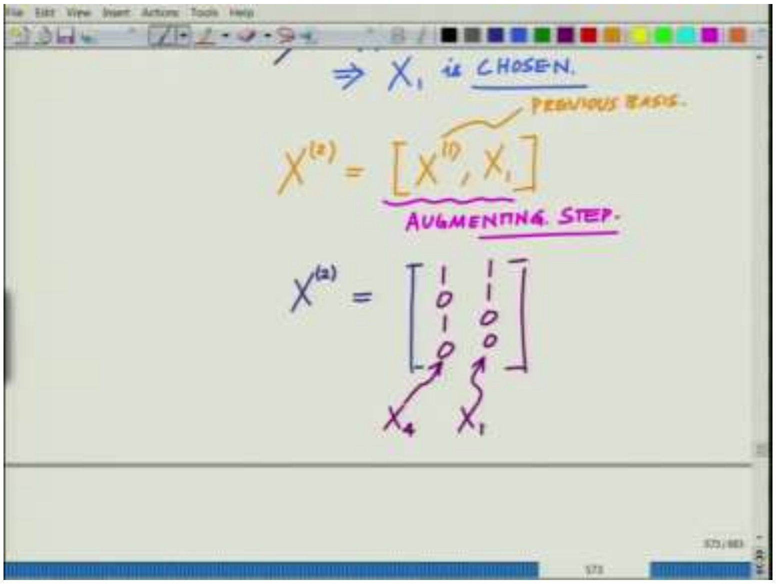
So, the i 2 is index, so this is the index, remember we are not interested in the value of the projection but we are interested in rather the index of the element that has the maximum projection, this means essentially the column one has the maximum projection.

(Refer Slide Time: 22:43)



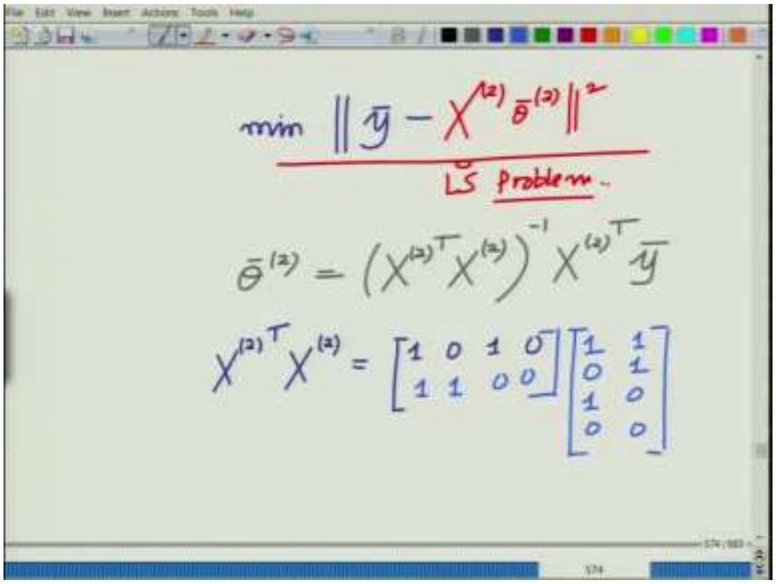
So, this implies that if you look at this i 2 equal to 1 implies that X 1 is, this implies that X 1 is chosen and which has the maximum projection on the residue and now what we do is, now we construct augment of the basis matrix. So, you construct the basis matrix. You take X the first and then argument it with the chosen column. So, this is the previous basis and this is essentially the, this is basically the augmenting step.

(Refer Slide Time: 23:43)



And now you will have, and therefore, now if you do this you will have X2 which is equal to, well already you had X2 equal to 1 0 1 0 that is the fourth column. Now we will add the first column that is 1 1 0 0, so these, remember this is your X4 column chosen in the first iteration. This is your X1 column chosen in the second iteration. Now once again we solve the least squares problem. Remember finding the best possible across approximation to y, norm of y bar minus X2 theta 2 whole square.

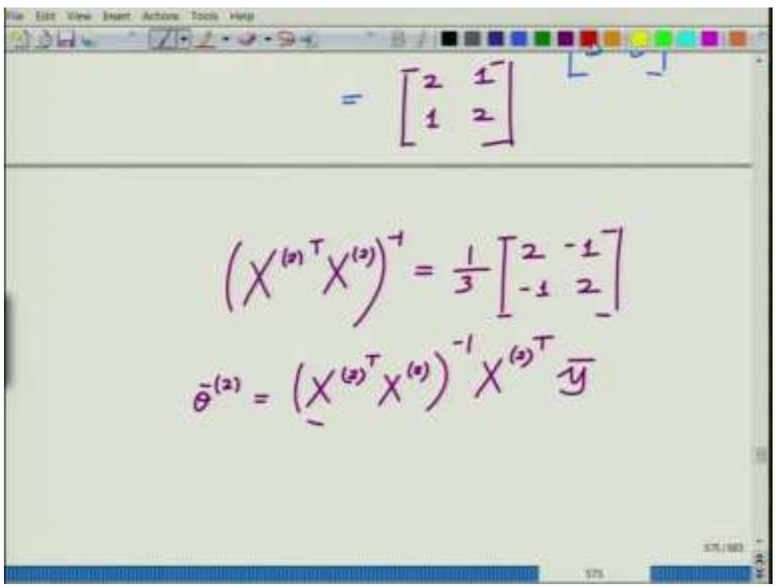
(Refer Slide Time: 24:27)

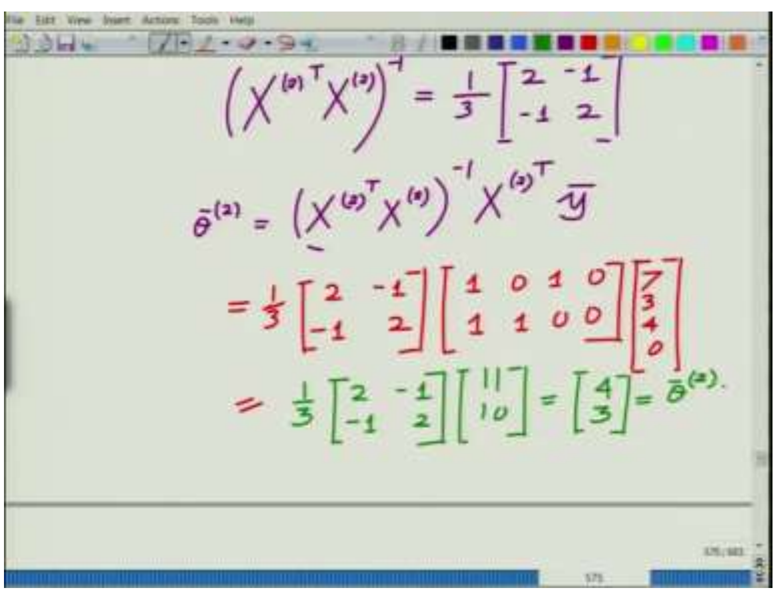


So, now once again we will solve the, merrily solve, I would say your least squares problem which we have, which you should be very comfortable with because remember as I told you this arises very frequently in all of linear algebra and the solution is if you remember and this is also something that would serve you well to remember at that tip, the top of your head X2 transpose X2 inverse times X2 times y bar.

Which now if I substitute X2 transpose, let us now substitute, now first let us evaluate this in two steps. Let us first, because it is a little complex. Let us employ X first, evaluate X2 transpose X2, that is going to be given as 1 0 1 0, 1 1 0 0, so this is 1 0 1 0, 1 1 0 0.

(Refer Slide Time: 25:57)

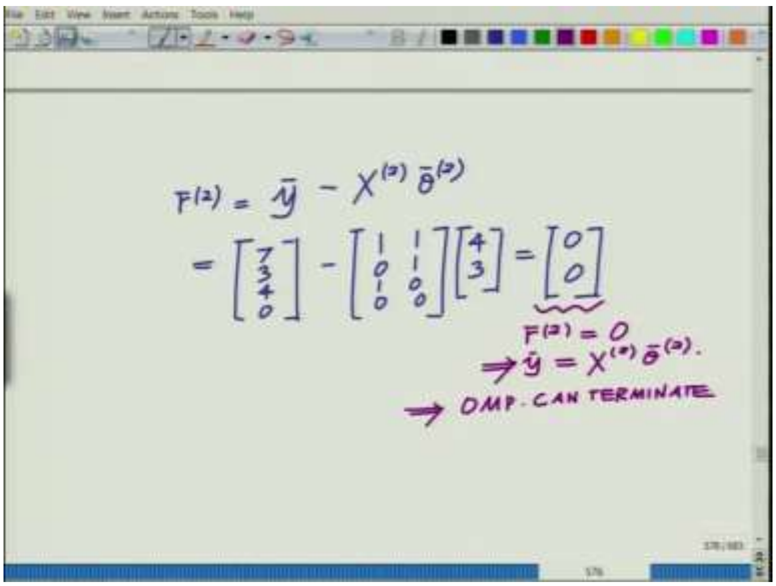




And therefore you can check, this is given as basically 2 1 1 2 and X2 transpose X2 inverse will therefore be 1 over the determinant, 1 over 3 times you have to exchange the diagonal elements negative of diagonal elements, this is the inverse. Now you have to multiply, so now you get theta bar 2 this is equal to X2 transpose X2 inverse times X2 transpose y bar. Which is equal to essentially, if you look at this 1 over 3, 2 minus 1, minus 1 2 times X2 transpose which is 1 0 1 0, 1 1 0 0 times y bar, which is essentially, remember this is the vector.

So, this you can evaluate this as 1 over 3, 2 minus 1, minus 1 2 times 1 0 1 0, 7 3 4 0 that is 7 plus 4 11. 1 1 0 0, that is 7 plus 3 10 and now if you evaluate this. This is 22 minus 10 12 divide by 3 that is 4 and this is 20 minus 11 that is 9 divide by 3 that is 3, so this is essentially your theta bar 2. And now once again to complete this iteration. So, we found X2 that is basis matrix, augmented basis matrix, we have to find theta bar 2, estimate of the non 0 coefficients of the parameter vector. Now find the residue to complete this iteration. What is the residue?

(Refer Slide Time: 28:23)

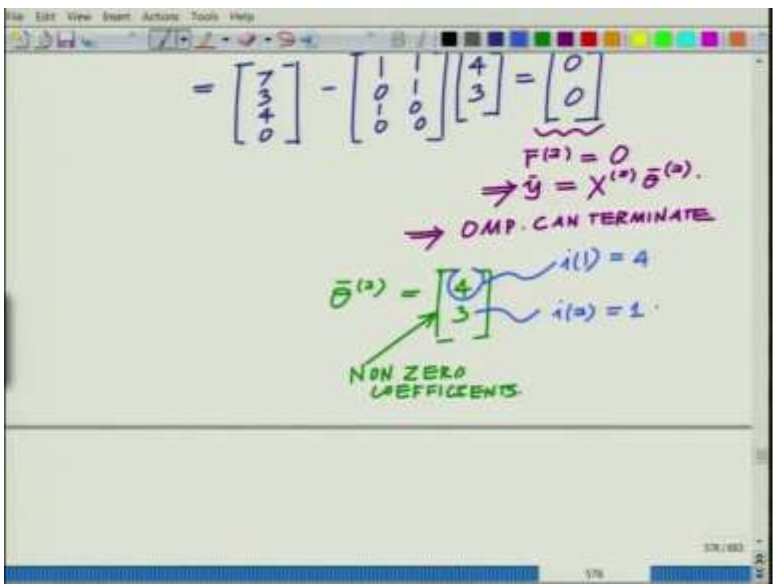


Now if you look at this residue, you will find something interesting. So, this is basically y bar minus X2 theta bar 2, which is 7 3 4 0 minus the matrix which is essentially what do we have over here. Which is 1 0 1 0 times 4 comma 3, which you can clearly see, this residue is equal to 0, which means essentially we have achieved y bar equals to X theta bar and therefore algorithm terminates. Remember we said one of the conditions is if residue is lower than a threshold, so residue is 0 which means now we are able to ideally approximate y bar.

Of course, this is the construed example and therefore, you can say residue equal 0, in practice you will never see residue actually equaling 0 because there will always be, so you can terminate this algorithm whenever the residue is below a certain acceptable threshold that depends of course on the noise variance. For instance if the noise is very high then the difference is going to be significant. If the noise is lower than the difference, the final residue can be very low. You can achieve a very good approximation.

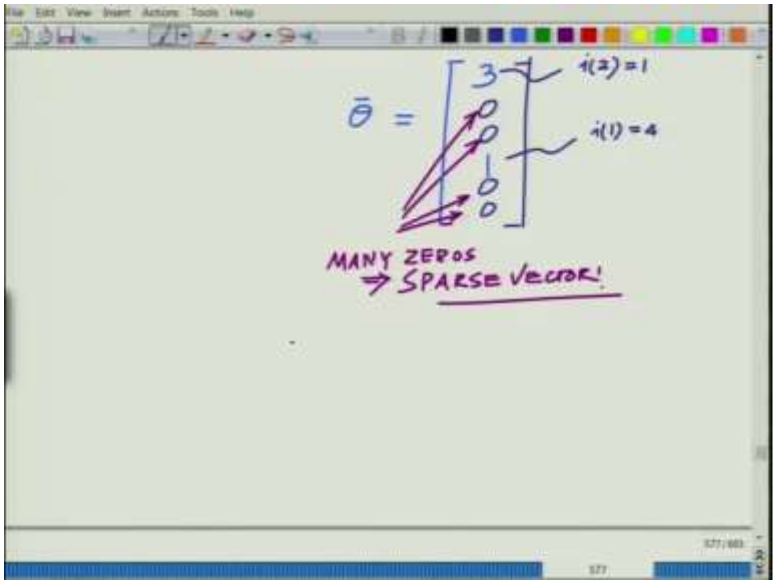
So, this is basically your r bar 2 which is equal to 0 which implies that y bar is exactly, y bar is, which implies basically y bar equals X2 comma theta bar 2 which implies that OMP can now terminate. So, the algorithm, OMP algorithm can now terminate and remember we have the non 0 coefficient.

(Refer Slide Time: 30:49)



Now look at this we still have one step left, remember theta bar equal to 2 which is basically if you look at it these are the coefficients, but these are the non 0 coefficients, these are not theta bar because theta bar will contain many 0's but these are the non 0 coefficients of, so how do we map them to theta bar? Now look at this, we have to look at the indices. Now this is corresponds to the basis column X1, column chosen in the first iteration, index chosen in the first iteration that is i 1 which is equal to 4. So, this is i 1 which is equal to 4 this corresponds to i 2 which is equal to 1.

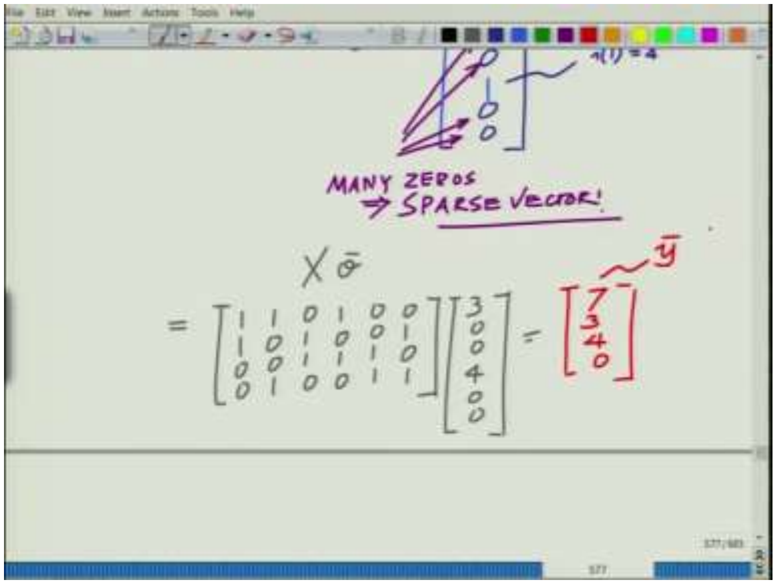
(Refer Slide Time: 31:41)



So, you will have the fourth entry, so you will have, in theta bar you will have the first entry as 3, fourth entry as 4 and the rest, remember this corresponds to your i2 which is index 1 and this corresponds to your i1 which is equal to 4, i2 which is equal to 1 and the rest of the entries are 0. So, this is essentially your theta bar.

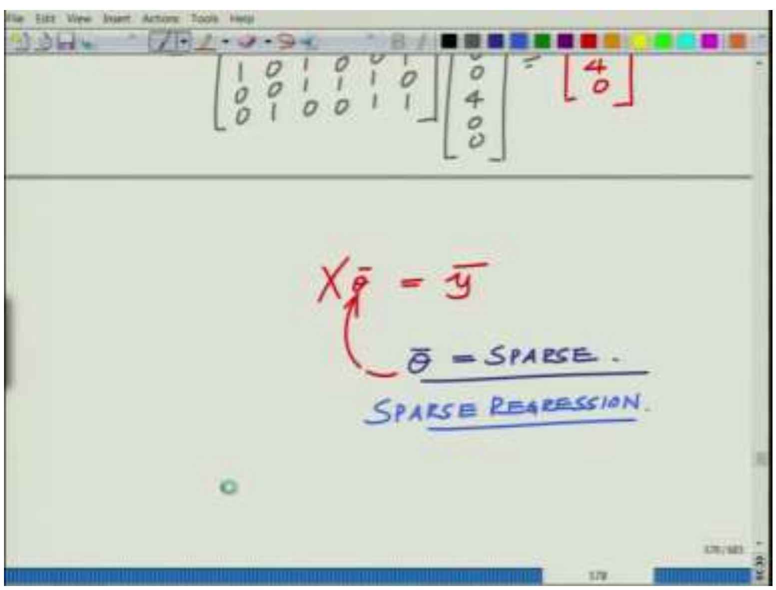
And therefore, you can clearly see theta bar is, there are many 0's in theta bar. Many 0's implies this is actually the sparse vector that we were talking about and now just to check you can actually check by substituting this in the original equation by looking at X times theta bar. Let us look at that also.

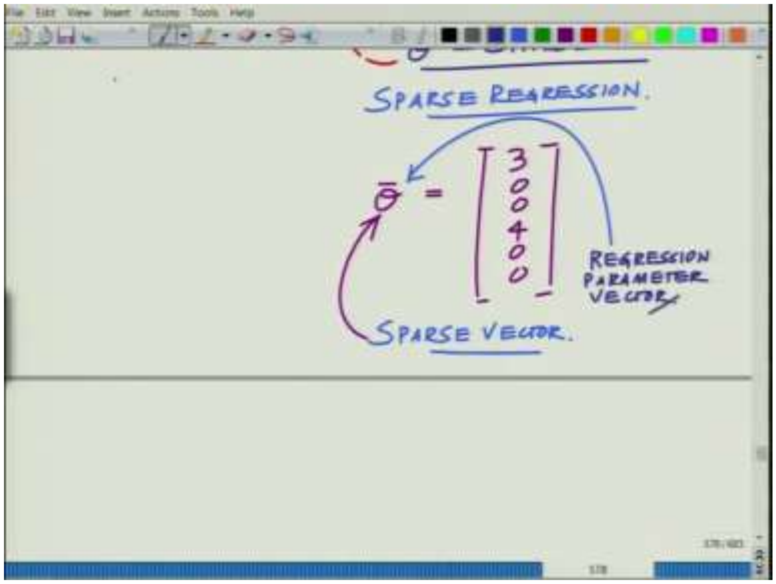
(Refer Slide Time: 33:07)



So, what is X times theta bar? This is equal to 1 1 0 0 let me write the original matrix 1 0 0 1, 0 1 1 0, 1 0 1 0, 0 0 1 1, 0 1 0 1 times theta bar which now is 3 0 0 4 0 0 and now when you multiply that, you can see what you get is, you get the 3 plus 4 7, 3 4 and then 0, which needless to say this is equal to y bar.

(Refer Slide Time: 34:10)





So, we have achieved a solution such that X theta bar equal to y bar, more importantly theta bar is sparse. So, this completes basically our sparse regression. All right and we have already seen that theta bar. Remember we have derived the vector theta bar which is basically if you look at this that is the vector 3 0 0 4 0 0 and this is basically what we are calling as our sparse vector. And this is also, remember this is our regression vector or regression parameter vector.

So, that essentially concludes this OMP algorithm and it is very simple algorithm but it is very powerful in the sense it clearly illustrates the step by step procedure that is required to solve this sparse regression problem and find the regression coefficients, that is basically the various thetas and what we have demonstrated is that this OMP algorithm indeed achieves or is able to find the sparse vector theta bar.

In which many thetas are 0, many values of theta i are 0 and only very few theta, in fact only two, there are only two non 0 values of theta. All right, so let us stop this example over, let us stop this module over here and we will continue with other aspects in the subsequent modules. Thank you very much.